



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO

Extração de relações semânticas binárias a partir de dados não estruturados em português

Trabalho de Conclusão de Curso

Ariel Ferreira Rodrigues



São Cristóvão – Sergipe

2019

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO

Ariel Ferreira Rodrigues

**Extração de relações semânticas binárias a partir de dados
não estruturados em português**

Trabalho de Conclusão de Curso submetido ao Departamento de Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador(a): Hendrik Teixeira Macedo

São Cristóvão – Sergipe

2019

Resumo

O presente trabalho avalia o uso de redes neurais convolucionais para extração de relações semânticas binárias de textos em português do Brasil e português de Portugal. Também é avaliado neste trabalho a influência da quantia de dimensões dos vetores de palavras nas macro-métricas (precisão, cobertura e métrica F1) resultantes. Para isso, foram avaliados quatro algoritmos de *word embeddings* diferentes (*Word2Vec*, *Wang2Vec*, *FastText*, *GloVe*), tendo sido os três primeiros treinados a partir das abordagens *SkipGram* e *CBoW*. Cada *word embedding* foi avaliado com os valores de dimensões 50, 300 e 1000. O conjunto de dados utilizado neste trabalho foi gerado utilizando supervisão distante e inclui mais de 90 mil relações semânticas e mais de 1 bilhão de tokens. Os experimentos realizados alcançaram precisão de 80% e cobertura de 72%, resultados superiores aos apresentados pelos autores do conjunto de dados adotado, também pôde-se perceber que a quantia de dimensões dos *word embeddings* não influencia de forma linear as macro-métricas resultantes, reforçando resultados já encontrados na literatura.

Palavras-chave: Processamento de linguagem natural, Extração de informações, Extração de relações, Redes neurais convolucionais.

Lista de ilustrações

Figura 1 – Arquitetura de uma Rede Neural Convolucional para classificação de sentimentos	16
Figura 2 – Arquitetura de uma Rede Neural Convolucional para extração de relações .	18
Figura 3 – Arquitetura de uma Rede Neural Convolucional para extração de relações .	24

Lista de tabelas

Tabela 1 – Caracterização estatística do conjunto de dados	21
Tabela 2 – Distribuição de sentenças em classes no conjunto de dados disponibilizado (BATISTA D.S.; FORTE, 2013)	21
Tabela 3 – Distribuição de sentenças em classes no conjunto de dados gerado neste trabalho	22
Tabela 4 – Classificação de relações utilizando o modelo de <i>word embeddings GloVe</i> .	30
Tabela 5 – Classificação de relações utilizando o modelo de <i>word embeddings Word2Vec CBoW</i>	31
Tabela 6 – Classificação de relações utilizando o modelo de <i>word embeddings Word2Vec Skip-Gram</i>	31
Tabela 7 – Classificação de relações utilizando o modelo de <i>word embeddings Wang2Vec CBoW</i>	32
Tabela 8 – Classificação de relações utilizando o modelo de <i>word embeddings Wang2Vec Skip-Gram</i>	32
Tabela 9 – Classificação de relações utilizando o modelo de <i>word embeddings FastText CBoW</i>	33
Tabela 10 – Classificação de relações utilizando o modelo de <i>word embeddings FastText Skip-Gram</i>	33
Tabela 11 – Comparação entre macrométricas dos trabalhos que utilizam o mesmo conjunto de dados	34
Tabela 12 – Comparação entre resultados por classes: Este trabalho - Wang2Vec Skip-Gram 1000 dimensões e (BATISTA D.S.; FORTE, 2013; BATISTA, 2016) .	34

Lista de abreviaturas e siglas

PLN	Processamento de Linguagem Natural)
EI	Extração de Informação
ER	Extração de Relações
ELI	<i>English Language Interpreter</i>
JASPER	<i>Journalist's ASsistant for Preparing Earning Reports</i>
DARPA	<i>Defense Advanced Research Projects Agency</i>
MUC	<i>Message Understanding Conferences</i>
ACE	<i>Automatic Content Extraction</i>
SVM	<i>Support Vector Machines</i>
LSA	<i>Latent Semantic Analysis</i>
HAL	<i>Hyperspace Analogue to Language</i>
GloVe	<i>Global Vectors</i>
CBoW	<i>Continuous Bag-of-Words</i>
CNN	<i>Convolutional Neural Network</i> (Rede neural convolucional)
NLCI	Núcleo Interinstitucional de Linguística Computacional
T_p	<i>True positive</i> (Positivos verdadeiros)
F_p	<i>False positive</i> (Falsos positivos)
F_n	<i>False negative</i> (Falsos negativos)
P	Precisão
C	Cobertura
ADAM	<i>Adaptative Moment Estimation</i>
kNN	<i>k-Nearest Neighbors</i> (K vizinhos mais próximos)

Sumário

1	Introdução	7
1.1	Definição do problema	8
1.2	Objetivo	8
1.2.1	Objetivo Geral	8
1.2.2	Objetivos Específicos	8
2	Revisão bibliográfica	9
2.1	Extração de informação	9
2.2	Extração de relações	10
2.2.1	Métodos baseados em conhecimento	11
2.2.2	Métodos supervisionados	11
2.2.3	Supervisão distante	11
2.3	Representação de palavras no espaço vetorial	13
2.3.1	Algoritmos para geração de vetores de palavras (<i>word embeddings</i>)	14
2.3.2	Estratégias de treinamento dos <i>word embeddings</i>	14
2.4	Redes Neurais Convolucionais	15
2.4.1	Convolução	16
2.4.2	<i>Pooling</i>	17
2.4.3	Uso de CNNs em tarefas de extração de relações	17
2.4.4	Extração de relações em português	18
3	Metodologia	20
3.1	Conjunto de dados	20
3.2	<i>Word embeddings</i>	23
3.3	Métricas de avaliação	23
3.4	Arquitetura da rede neural convolucional	24
3.4.1	A camada de entrada	25
3.4.2	A Camada convolucional	26
3.4.3	A camada de pooling	26
3.4.4	A camada de classificação	27
3.5	Experimentos e reprodutibilidade	27
4	Experimentos, resultados e discussão	29
5	Conclusão	36
	Referências	38

1

Introdução

É interesse de todos os membros de uma democracia avaliar e ponderar acerca das propostas dos candidatos ao governo. Pensando nisso, foi determinado pela Lei nº 9.504, de 29 de setembro de 2009, que candidatos à Chefia do Executivo no estado democrático de direito brasileiro devem apresentar à Justiça Eleitoral, juntamente com o seu requerimento de registro de candidatura, uma via impressa e outra digitalizada de sua plataforma ou plano de governo ¹.

Apesar disto, não há um padrão estabelecido para a formatação ou apresentação das propostas dos candidatos nos planos de governo. Desta forma, analisar e comparar os planos de governo de todos os candidatos aos cargos de Chefia do Executivo, apesar de ser uma tarefa de grande importância, é exaustiva para ser realizada de forma manual.

Automatizar o processo de extração ou até classificação das propostas contidas no plano de governo dos candidatos tornaria o processo de análise e comparação dos planos de governo mais simples. O que também contribuiria para a qualidade da informação que chega ao eleitor e, consequentemente, a qualidade da democracia do nosso país.

A área de pesquisa na computação que trabalha a extração de informações, a partir de dados não estruturados, surge ainda durante a década de 1950, a partir da intersecção entre a inteligência artificial e a linguística, e fica conhecida como Processamento de Linguagem Natural (PLN). Atualmente as pesquisas em PLN abordam diversos temas presentes em nosso dia-a-dia como por exemplo assistentes virtuais e tradução automática em tempo real.

O problema de extrair dados estruturados a partir de textos é tratado por um campo específico da PLN chamado Extração de Informação (EI). Este trabalho tem como foco principal o desenvolvimento de um *pipeline* para extração de relações semânticas a partir de textos utilizando estudos acerca do estágio da área de EI conhecido como Extração de Relações (ER).

¹ <<http://www.tse.jus.br/legislacao/codigo-eleitoral/lei-das-eleicoes/lei-das-eleicoes-lei-nb0-9.504-de-30-de-setembro-de-1997>>

Após o desenvolvimento das ferramentas para extração de relações semânticas, os resultados deste trabalho serão aplicados a um trabalho maior que envolve o processamento dos planos de governo com objetivo de realizar a identificação e extração de promessas dos candidatos nas eleições brasileiras.

1.1 Definição do problema

A extração de relações semânticas é uma etapa intermediária comum em diversas tarefas na área de processamento de linguagem natural (PLN). Apesar de existirem uma variedade de abordagens para realizar a tarefa de extrair relações semânticas a partir de textos em inglês, infelizmente, nem todos os idiomas contam com esta variedade de recursos disponíveis. Para o idioma português (do Brasil e Portugal), por exemplo, recursos como corpora rotulados e estratégias bem definidas para tratar o problema da extração de relações neste idioma ainda são escassos se comparado com a quantidade de recursos disponíveis para o idioma inglês. Portanto, para mudar esse cenário, é de grande importância a produção de trabalhos que avaliem o desempenho de abordagens para extração de relações já existentes utilizando textos no idioma português.

1.2 Objetivo

1.2.1 Objetivo Geral

Apresentar e avaliar um modelo de rede neural convolucional (*CNN*) para a tarefa de extração de relações semânticas binárias em português;

1.2.2 Objetivos Específicos

- Avaliar o modelo proposto por (MELO, 2018) para a extração de relações semânticas binárias a partir da representação vetorial de sentenças em português utilizando redes neurais convolucionais;
- Avaliar a influência do número de dimensões dos *word embeddings* em tarefas de extração de relações semânticas;
- Avaliar o uso de um conjunto de dados rotulado para a tarefa de extração de relações em português gerado utilizando supervisão distante;

2

Revisão bibliográfica

2.1 Extração de informação

A Extração de Informações tem como objetivo a extração de conhecimento a partir de documentos escritos em linguagem natural, podendo ser descrita como “o processo de extrair informações estruturadas a partir de dados não estruturados” (JURAFSKY D.; MARTIN, 2009).

Os primeiros registros de pesquisas na área de extração de informação se passam na década de 1970; dentre os de maior relevância podemos citar *ELI* (*English Language Interpreter*) (RIESBECK C. K.; SCHANK, 1976) capaz de produzir representação estruturada de informação semântica em histórias, *Foul-UP* (GRANGER, 1977) capaz de construir definições específicas, levando em consideração o domínio, para palavras desconhecidas e o trabalho de (LEHNERT, 1977), que combina ideias de processamento de informação conceitual (*conceptual information processing*) (SCHANK, 1975), memória humana e computação para propor uma teoria de perguntas e respostas (*question answering*). Lehnert implementou a sua teoria em dois sistemas de histórias.

Apesar desse novo campo ter atraído a atenção de pesquisadores da área de inteligência artificial desde a década de 1970, os primeiros sistemas comerciais da área de extração de informações só começam a surgir a partir do final dos anos 1980, com o *JASPER* (*Journalist's ASsistant for Preparing Earning Reports*) (ANDERSEN, 1992). *JASPER* foi construído com objetivo de fornecer notícias financeiras em tempo real para os comerciantes e utilizava um conjunto de modelos e heurísticas feitos à mão para produzir fatos a partir de notícias recentes. *JASPER* era capaz de automatizar a rotina tediosa com acurácia variando entre 61% e 96% em diversas atividades de extração rotuladas.

As pesquisas na área de Extração de Informação ganham ainda mais fôlego no início dos anos 1990 quando a *DARPA* (*Defense Advanced Research Projects Agency*) deu início à uma série de competições e conferências com objetivo de estimular o desenvolvimento de novas técnicas

para extração de informações para uso militar conhecidas como *MUC* (*Message Understanding Conferences*) (MESSAGE..., 1996). Desde então, as pesquisas nessa área não pararam de crescer. O objetivo de cada conferência era executar uma tarefa de extração de informações bem definida de forma a surgirem novas soluções que impulsionassem o estado-da-arte.

Diversas tarefas independentes e importantes para a atividade e extração de informação foram definidas no *MUC* (MESSAGE..., 1996), como por exemplo o reconhecimento de entidades nomeadas e a extração de relações. A decomposição dos sistemas de extração de informação em tarefas de menor complexidade, como a extração de relações, tem como principal vantagem transformar tarefas grandes e complexas em tarefas menores, mais simples e intercambiáveis. Essas tarefas menores servem como suporte para a construção de sistemas mais complexos de extração de informações. Além disso, muita dessas tarefas podem ser reduzidas a problemas de classificação, dispondo assim de muitos anos de conhecimento adquirido por pesquisadores da área de aprendizagem de máquina para ser utilizado no processamento de linguagem natural.

2.2 Extração de relações

A extração de relações é uma subárea da extração de informações que consiste na identificação de relações semânticas entre entidades em documentos escritos em linguagem natural. Desta forma, uma tarefa de extração de relações pode ser descrita como: dada uma sentença S com pelo menos um par de entidades e_1 e e_2 , devem ser identificadas relações como tuplas da forma $(R, e_1, e_2, \dots, e_n)$, onde cada e_i é uma entidade distinta, mas não necessariamente única, e R é o tipo da relação. Por exemplo, a partir da sentença: "Maria disse que João trabalha na UFS" pode ser extraída a seguinte relação: `trabalha_na(João, Ufs)`

Parte do trabalho realizado nas tarefas de extração de relações é baseado na tarefa definida a partir do programa *ACE* (*Automatic Content Extraction*) (DODDINGTON, 2004). O *ACE* focou na extração de relações binárias, ou seja, aquelas que envolvem apenas duas entidades. Alguns exemplos de tipos de relações definidas no *ACE* são:

- *Near*: quando uma entidade está fisicamente próxima de outra entidade;
- *Social*: quando uma pessoa é membro da família de outra pessoa, ex. parentes, primos, cônjuge;
- *Role*: o papel desempenhado por uma pessoa em uma organização, ex. gerente, membro, dono, fundador.

Diversas abordagens podem ser utilizadas para a extração de relações semânticas, nesta seção discutiremos as abordagens utilizando conhecimento especialista, métodos supervisionados e métodos que utilizam supervisão distante.

2.2.1 Métodos baseados em conhecimento

A extração de relações utilizando métodos baseados em conhecimento se mostra presente principalmente em tarefas de domínio específico, onde há um conjunto de relações pré determinado a serem identificadas. Sistemas que utilizam esta abordagem contam com casamento de padrões utilizando regras feitas manualmente para domínios específicos (RILOFF E.; JONES, 1999), ou utilizando padrões genéricos para extração de relações em domínios distintos (HEARST, 1992).

Os resultados obtidos com esta abordagem normalmente apresentam alta precisão e baixa cobertura. O fator que mais contribui para estes resultados é a geração manual de padrões para extração de relações, comumente os padrões gerados são específicos o suficiente para gerarem poucos erros. Entretanto, neste caso, muitas relações são eliminadas por não encaixarem nos padrões utilizados (HUMPHREYS K., 1998).

2.2.2 Métodos supervisionados

Na abordagem utilizando métodos supervisionados os extratores de relações utilizam aprendizagem de máquina e são treinados utilizando grandes conjuntos de dados rotulados. A abordagem mais comum utilizando este método trata a tarefa como um problema de classificação. Dada uma sentença onde há a ocorrência de pelo menos um par de entidades, a tarefa consiste em classificar a relação ou as relações entre entidades utilizando tipos de relações conhecidos. O principal passo dessa abordagem é a definição das características nas quais se baseará a classificação.

Entretanto, esta abordagem exige uma grande quantia de dados rotulados para treinamento. Desta forma, a geração de dados de treinamento de forma supervisionada se torna inviável, tendo em vista a grande quantia de esforço humano necessário para gerar grandes quantias de dados de treinamento de forma manual. Assim, surgem alternativas como a geração de dados de treinamento de forma fracamente supervisionada, que pode ocorrer a partir do treinamento do sistema a partir de um pequeno conjunto de dados rotulados e, de forma interativa, são encontradas instâncias de novas relações assim como novos padrões de extração (*bootstrapping*); ou ainda a supervisão distante que envolve o uso de instâncias de relações conhecidas a partir de bases de conhecimento como WordNet-BR como dados de treinamento do sistema.

2.2.3 Supervisão distante

Com o objetivo de desenvolver alternativas à construção manual de conjuntos de dados rotulados surgem abordagens a partir de supervisão distante. Estas consistem no uso de fontes de conhecimento externo para superar as limitações das metodologias de criação de conjuntos de dados mais tradicionais. A supervisão distante se torna particularmente interessante em domínios onde há grandes quantias de dados não rotulados e pelo menos uma fonte de dados

rotulados estruturados. Neste cenário, pode ser aplicada uma estratégia bastante comum de supervisão distante que consiste em associar, a partir de heurísticas, a fonte de dados estruturados e rotulados com os dados não rotulados (MINTZ M.; BILLS, 1992; NIGAM K.; MCCALLUM, 1998; CRAVEN M.; KUMLIEN, 1999; BANKO M.; CAFARELLA, 2007; MINTZ M.; BILLS, 2009; NGUYEN; MOSCHITTI, 2011; SCHMITZ M.; BART, 2012; ZENG D.; LIU, 2014; XIANG Y.; CHEN, 2016; MELO, 2018).

A hipótese por trás da abordagem de supervisão distante é a de que se uma relação entre duas entidades existe em uma determinada base de conhecimento, então há uma grande probabilidade de que dada uma sentença, do mesmo domínio da base de conhecimento, que mencione as mesmas duas entidades a relação presente é a mesma que a mapeada na base de conhecimento.

Por exemplo, dada a relação

trabalha_na(Maria, UFS)

e dadas as sentenças

*"... e então João disse que **Maria** chegou na **UFS**, onde trabalha."*

*"... já **Maria**, funcionária da **Universidade Federal de Sergipe (UFS)**, afirma ..."*

Aplica-se o mesmo procedimento que é aplicado na aprendizagem supervisionada, são extraídas características dos textos e os modelos são treinados.

Um exemplo prático de abordagem utilizando supervisão distante é dado por (MINTZ M.; BILLS, 2009). Nesse trabalho, foi utilizada a Freebase (BOLLACKER K.; EVANS, 2008), uma base de dados estruturados de informação semântica, como uma forma de construir dados rotulados. Para cada par de entidades (e_1 , e_2) referenciados no Freebase, foram buscadas sentenças que continham as entidades em um corpus pré estabelecido. Posteriormente, essas sentenças foram utilizadas como dados de treinamento de um extrator de relações baseado em um classificador de máxima entropia que combinava atributos léxicos e sintáticos. Sendo utilizado desta forma, um conjunto de dados com 10.000 sentenças e 102 tipos de relações diferentes foi gerado a partir da supervisão distante para treinar um classificador supervisionado.

(NGUYEN; MOSCHITTI, 2011), por sua vez, utiliza dados de treinamento da *Yet Another Great Ontology (YAGO)* (SUCHANEK F. M.; KASNECI; WEIKUM, 2007) e sentenças de documentos da *Wikipedia* mencionados no *Freebase* para treinar um extrator de relações semânticas baseado utilizando métodos de *kernel*. Nesse trabalho foram desenvolvidos 52 extratores e a métrica F1 média foi de 74,3%.

A principal desvantagem das abordagens utilizando supervisão distante é o ruído que pode estar presente nas sentenças. Isto ocorre pois nem todas as sentenças que contêm as duas

entidades, expressa a mesma relação semântica entre as entidades que a base de conhecimento. Por exemplo, considerando a relação citada acima `trabalha_na`(Maria, UFS) a seguinte sentença poderia ser selecionada para o conjunto de treinamento

*"Por causa do calor extremo, **Maria** decidiu não ir à UFS."*

Entretanto, esta sentença não expressa a relação **trabalha_na** e seria considerada um ruído no conjunto de treinamento. Diversas abordagens são propostas para diminuir a quantidade de relações ruidosas presentes no conjunto de treinamento, dentre estas, destacam-se a *at-least-one* (HOFFMANN R.; ZHANG; WELD, 2011; RIEDEL S.; YAO; MCCALLUM, 2010) e *Patterns Correlations* (TAKAMATSU S.; SATO; NAKAGAWA, 2012).

2.3 Representação de palavras no espaço vetorial

A representação de palavras como símbolos atômicos discretos tem funcionado muito bem para nós humanos, mas essa representação não expressa características relevantes para os computadores. Isso acontece pois essa forma de representar as palavras não inclui características (*features*) variadas acerca da informação que está sendo representada. Uma outra forma de representação de palavras é através dos *word embeddings* (BENGIO Y.; DUCHARME, 2003; COLLOBERT R.; WESTON, 2011; MIKOLOV T.; CHEN, 2013; LING W.; DYER, 2015; LAI S.; XU, 2015; HARTMANN N.; FONSECA, 2017). A representação vetorial permite mapear palavras em vetores de dimensionalidade reduzida, onde cada dimensão do vetor carrega um valor contínuo correspondente a uma característica da palavra. Esta forma de representação inclui diversas características sintáticas e semânticas da palavra que está sendo representada (MIKOLOV T.; CHEN, 2013).

Os diferentes algoritmos que tem sido desenvolvidos para gerarem *embeddings* podem ser grosseiramente classificados em duas grandes famílias de métodos (BARONI M.; DINU, 2014).

A primeira é composta de métodos que trabalham com a co-ocorrência de matrizes de palavras, como o *Latent Semantic Analysis (LSA)* (DUMAIS S. T.; FURNAS, 1988), o *Hyperspace Analogue to Language (HAL)* (LUND K.; BURGESS, 1996) e o *Global Vectors (GloVe)* (PENNINGTON J.; SOCHER, 2014).

A segunda é composta dos métodos preditivos, aqueles que tentam prever as palavras vizinhas dada uma ou mais palavras, como por exemplo o *Word2Vec* (MIKOLOV T.; CHEN, 2013). Apesar disto, os experimentos realizados por (BARONI M.; DINU, 2014) indicam que os modelos baseados em predição tendem a superar em performance e eficiência computacional os modelos que trabalham com co-ocorrência de matrizes de palavras.

2.3.1 Algoritmos para geração de vetores de palavras (*word embeddings*)

Dentre a variedade de modelos de *word embeddings* disponíveis, foram adotados quatro modelos para os experimentos realizados neste trabalho. Nesta subseção são descritos de forma breve os algoritmos dos modelos adotados para este trabalho.

- *GloVe (Global Vectors)*: algoritmo utilizado para gerar modelos de *word embeddings* proposto por (PENNINGTON J.; SOCHER, 2014) que consiste na construção de uma matriz de co-ocorrência \mathbf{M} a partir das palavras de contexto. Cada elemento M_{ij} na matriz de co-ocorrência representa a probabilidade da palavra i estar próxima da palavra j .
- *Word2Vec*: algoritmo amplamente utilizado para gerar modelos de *word embeddings* proposto por (MIKOLOV T.; CHEN, 2013). Os modelos gerados consistem apenas de uma única matriz de pesos e são capazes de capturar informações semânticas dos dados processados. Além disso, o treinamento de modelos com *Word2Vec* tem complexidade *log-linear*.
- *Wang2Vec*: algoritmo utilizado para gerar modelos de *word embeddings* proposto por (LING W.; DYER, 2015). O *Wang2Vec* surge como uma modificação do *Word2Vec* com objetivo de levar em consideração a falta de ordem de palavras. As modificações propostas pelo *Wang2vec* têm como objetivo tornar os *embeddings* melhores em capturar o comportamento sintático das palavras.
- *FastText*: algoritmo utilizado para gerar modelos de *word embeddings* proposto por (BOJANOWSKI P.; GRAVE, 2016). No *FastText* os *embeddings* são associados com *n-grams* de caracteres e as palavras são representadas como a composição dessa representação de caracteres. Desta forma, este método tenta capturar e adicionar nos *word embeddings* informação morfológica das palavras.

2.3.2 Estratégias de treinamento dos *word embeddings*

Os *embeddings* gerados a partir dos algoritmos do *Word2Vec*, *Wang2Vec* e *FastText* podem ser treinados utilizando duas estratégias distintas:

Skip-gram

O modelo *Skip-gram* tem como objetivo a partir de um corpus de palavras w e seu contexto c prever o contexto de cada palavra w_i pertencente a w . Para atingir este objetivo considera-se a probabilidade condicional $p(c|w; \theta)$, onde deve ser determinado o parâmetro θ que maximiza esta probabilidade.

$$\operatorname{argmax}_{\theta} \prod_{(w,c) \in D} p(c|w; \theta) \quad (2.1)$$

onde D é o conjunto de todas as palavras e seus respectivos contextos extraídos do corpus disponível.

Desta forma, a entrada do modelo é cada palavra $w_i \in w$ e o seu respectivo contexto $C(w) = w_{i-k}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+k}$, sendo k o comprimento da janela determinado. O modelo treinado a partir desta entrada terá determinado a melhor constante θ tal que a probabilidade condicional $p(c|w; \theta)$ seja a máxima, conforme apresentado na equação 2.1.

Continuous Bag-of-Words (CBoW)

O procedimento realizado no modelo *CBoW* é bastante semelhante ao procedimento realizado no modelo *Skip-Gram*. A diferença mais notória entre estes modelos é que em vez de tentar maximizar a classificação do contexto dada uma determinada palavra, tenta-se prever qual a palavra dado o seu determinado contexto. Desta forma, o modelo tem como entrada um determinado contexto $C(w) = w_{i-k}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+k}$, para uma janela de palavras de tamanho k , e a saída esperada é a palavra w_i com maior probabilidade de pertencer ao contexto $C(w)$.

2.4 Redes Neurais Convolucionais

Originalmente desenvolvidas para tarefas de visão computacional, as redes neurais convolucionais (*Convolutional Neural Network, CNN*) também têm se mostrado bastante eficazes em tarefas de processamento de linguagem natural, apresentando resultados promissores em diversas tarefas de processamento de linguagem natural, como por exemplo Reconhecimento de entidades nomeadas, Extração de relações, Compreensão de linguagem natural e Classificação de textos (COLLOBERT R.; WESTON, 2011; ZHANG X.; LECUN, 2015; LAI S.; XU, 2015; MELO, 2018). As *CNNs* surgem como estratégia para emular o mecanismo básico do córtex visual animal e utilizam camadas com filtros convolucionais que são aplicados a características locais (LECUN Y.; BOTTOU, 1998). Os pesos nas redes neurais convolucionais são compartilhados entre todos os neurônios, sendo desta forma reduzido o número total de pesos treináveis e, portanto, introduzindo esparsidade na rede.

Para aplicar *CNNs* em documentos de texto, normalmente de alta dimensionalidade e de tamanhos variados, a informação contida nesses textos precisa antes ser convertida para uma representação de menor dimensionalidade. Para isso, são utilizadas representações vetoriais de palavras, como os *word embeddings* ou vetores esparsos como *1-hot*. Nesta representação, uma frase com n palavras e um embedding com m dimensões resultaria em uma matriz $n \cdot m$ representando a informação.

Desta forma, as camadas da *CNN* convertem pequenas regiões dos dados de entrada (ex. palavras de um documento de texto) em vetores de características para as camadas mais internas. Gerando nesse processo uma saída de menor dimensionalidade mas que preserva a

estrutura preditiva dos dados de entrada. Usualmente a entrada de uma *CNN* é dada pelas palavras codificadas como *embeddings*, camadas convolucionais com diversos filtros, camadas de *pooling* e, por último, um classificador de *softmax* (MELO, 2018; SILVA R. P. DA; SANTOS, 2018; KIM, 2014; KALCHBRENNER N.; GREFFENSTETTE, 2014; JOHNSON R.; ZHANG, 2014).

Na Figura 1 é apresentada um exemplo didático de uma *CNN* simples para extração de sentimentos a partir de sentenças. A entrada do modelo é dada pelo conjunto de palavras que compõem a sentença codificados em forma vetorial. Na camada seguinte é são extraídas características destas palavras a partir da convolução das mesmas com o filtro determinado pelo projetista da rede. O resultado obtido a partir da convolução tem sua dimensionalidade reduzida após aplicação do *max-pooling* e, por fim, é realizada uma regularização para evitar *overfitting* utilizando *dropout*. O resultado da rede é dado a partir aplicação do algoritmo de *softmax* no vetor resultante.

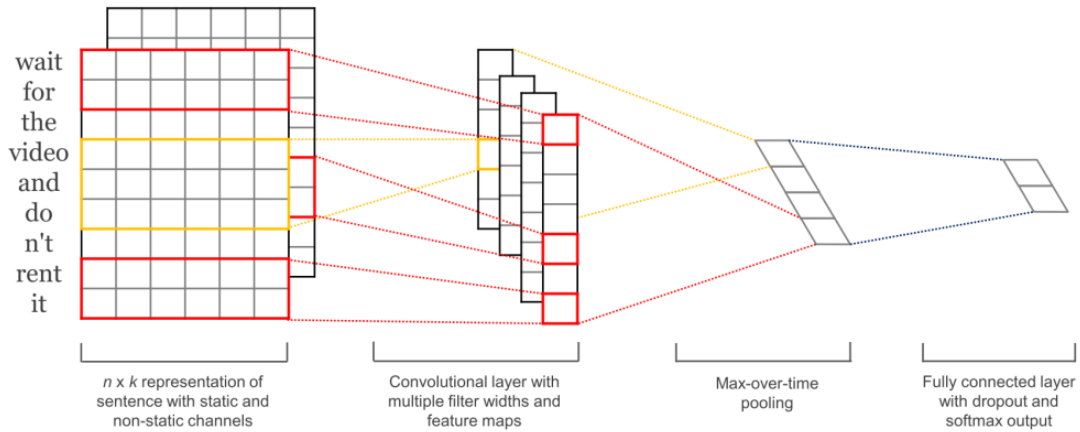


Figura 1 – Arquitetura de uma Rede Neural Convolucional para classificação de sentimentos

Fonte: (MELO, 2018)

2.4.1 Convolução

Convolução é uma operação linear matemática (denotada por $*$) que mede a soma do produto de duas funções ao longo da região subentendida pela superposição delas em função do deslocamento existente entre as mesmas. A operação de convolução no domínio discreto é dada por

$$(X * F)(i, j) = \sum_{a=0}^{k-1} \sum_{b=0}^{k-1} X(i-a, j-b) \cdot F(a, b) \quad (2.2)$$

Se \mathbf{X} for a matriz de entrada e \mathbf{F} um filtro (*kernel*) de dimensões $k \times k$, \mathbf{X} convolvido com \mathbf{F} pode ser entendido como a correlação cruzada entre \mathbf{X} e \mathbf{F} estando \mathbf{F} "invertido" horizontalmente e verticalmente.

Assim, a saída de cada camada da rede neural convolucional pode ser representado por

$$o_{i,j}^l = f(x_{i,j}^l) = f\left(\sum_{i'} \sum_{j'} w_{i',j'}^l \cdot o_{i-1,j-1}^{l-1} + b^l\right) \quad (2.3)$$

Sendo $l \in [1, L]$ a camada da *CNN*, $w_{i,j}^l$ o vetor de pesos que conecta o neurônio da camada l com neurônios da camada $l + 1$, b_l o escalar de polarização (*bias*) da camada l , $x_{i,j}^l$ o vetor de entrada convolvido na camada l somado ao escalar de polarização e $f(\cdot)$ a função de ativação.

2.4.2 Pooling

As *CNNs* podem contar ainda com uma camada responsável por reduzir progressivamente o tamanho espacial da representação, reduzindo assim o número de parâmetros e, portanto, controlando o *overfit*. Esta camada é chamada de *pooling* e tem como principal função reduzir a dimensionalidade da representação a um tamanho fixo, nenhum aprendizado ocorre nesta camada.

2.4.3 Uso de CNNs em tarefas de extração de relações

Em tarefas de classificação de relações, a rede neural convolucional recebe como entrada além das palavras representadas de forma vetorial, as posições relativas das palavras em relação às entidades participantes da relação semântica em questão. O pressuposto por trás desta metodologia é de que as posições das entidades ajuda a rede a ponderar sobre o quão próxima está cada palavra das entidades, sendo palavras mais próximas das entidades mais capazes de conter informações úteis sobre a classe da relação.

Estudos recentes indicam que a abordagem tem apresentado bons resultados em tarefas de classificação de texto com entrada de dimensões maiores, tornando-a uma linha de base padrão para novas arquiteturas de classificação de texto (MELO, 2018; ZENG D.; LIU, 2014; SANTOS C. N. D.; XIANG, 2015; NGUYEN T. H.; GRISHMAN, 2015; XU K.; FENG, 2015; ZENG D.; LIU, 2015).

Um exemplo desta abordagem é apresentado na 2.4.3, proposta em (NGUYEN T. H.; GRISHMAN, 2015). A camada de entrada desta rede neural recebe as sentenças codificadas em *word embeddings* e os respectivos *position embeddings*. Em seguida, estes dados seguem para a camada convolucional, onde são extraídas as características a partir da convolução com um filtro de tamanho variável, o principal objetivo desta camada é extrair as características acerca dos *n-grams*. As características extraídas a partir da convolução são encaminhadas à camada de *pooling* que extrai as melhores características e reduz a dimensionalidade dos resultados e em seguida a uma camada de regressão logística que realiza a classificação.

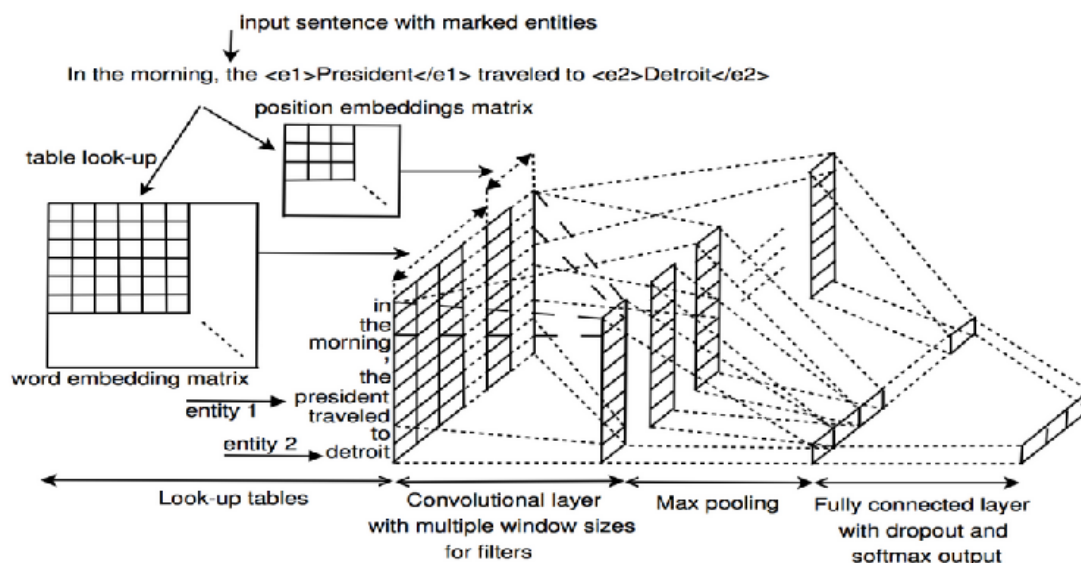


Figura 2 – Arquitetura de uma Rede Neural Convolucional para extração de relações

Fonte: (NGUYEN T. H.; GRISHMAN, 2015)

2.4.4 Extração de relações em português

A extração de relações em português do brasil e português do portugal, em comparação com a extração de relações a partir de textos no idioma inglês, é um tópico de pesquisa que ainda dispõe de poucos estudos.

Apesar da pequena quantia de estudos, diversas abordagens foram utilizadas para realizar a extração de relações em português. As abordagens utilizadas podem ser classificadas de forma grosseira em duas categorias: abordagens baseadas em regras (BICK, 2007; FREITAS M. C. DE; QUENTAL, 2007; OLIVEIRA H. G.; SANTOS, 2010; CARDOSO, 2008; PEREIRA V.; PINHEIRO, 2015; CRUZ G. A.; WEITZEL, 2018) e aprendizagem de máquina (ALVA-MANCHEGO F. E.; ROSA, 2012; SEQUEIRA J.; GONÇALVES, 2012; FONSECA E. R.; ROSA, 2012; FONSECA E. R.; ROSA, 2013; BATISTA D.S.; FORTE, 2013; HARTMANN N. S.; DURAN, 2017; CRUZ G. A.; WEITZEL, 2018).

A maioria dos trabalhos mais recentes que propõem uma estratégia para extração de relações em português utiliza aprendizagem de máquina, na maioria das vezes empregando técnicas de aprendizagem supervisionada. Um dos principais problemas enfrentados por trabalhos que optam por esta abordagem é a escassez de dados rotulados para extração de relações em português (ALVA-MANCHEGO F. E.; ROSA, 2012; FONSECA E. R.; ROSA, 2012; CRUZ G. A.; WEITZEL, 2018).

Algumas abordagens são propostas para superar a escassez de dados rotulados para extração de relações em português. (CRUZ G. A.; WEITZEL, 2018), por exemplo, utiliza o Re-Verb (FADER A.; SODERLAND, 2011) para classificar as relações do conjunto de treinamento. Uma outra proposta é dada por (BATISTA D.S.; FORTE, 2013) e consiste no uso de supervisão

distante para produzir um *corpus* utilizando sentenças do *Wikipedia* e relações do *DBPedia*.

3

Metodologia

A abordagem adotada neste trabalho é constituída por uma rede neural convolucional com apenas uma camada e foi proposta por (MELO, 2018). Em seu trabalho, Melo propõe uma abordagem utilizando redes neurais convolucionais para extração de relações a partir de textos em inglês. Neste trabalho optou-se por replicar a arquitetura proposta por Melo, realizando alterações apenas na camada de entrada para permitir a reprodutibilidade do seu trabalho no idioma português.

Desta forma, o método do trabalho seguiu as seguintes etapas:

1. Seleção do conjunto de dados para treinamento e teste do classificador de relações;
2. Seleção dos *word embeddings*;
3. Definição das métricas de avaliação adotadas para este trabalho;
4. Aplicação da arquitetura, treinamento e avaliação da rede neural convolucional proposta por (MELO, 2018);

3.1 Conjunto de dados

O conjunto de dados escolhido para a realização deste trabalho é composto por frases extraídas automaticamente da *Wikipédia* que expressam relações entre entidades extraídas da *DBPedia* e, portanto, gerado por supervisão distante por (BATISTA D.S.; FORTE, 2013). Este conjunto de dados foi escolhido por ser o maior conjunto de dados rotulados para extração de relações em português até o momento do início deste trabalho. O conjunto de dados é composto por 97.988 sentenças, 2.186.445 palavras, 10 classes e 71.119 entidades únicas. A caracterização estatística completa pode ser conferida na Tabela 1.

Tabela 1 – Caracterização estatística do conjunto de dados

	Total
Sentenças	97.988
Palavras	2.186.445
Classes	10
Instâncias	89.601
Entidades únicas	71.119
Média palavras/frase	22,43
Desvio Padrão palavras/frase	11,39
Média instâncias/classe	3.960,1
Desvio Padrão instâncias/classe	14.172,38

Tabela 2 – Distribuição de sentenças em classes no conjunto de dados disponibilizado (BATISTA D.S.; FORTE, 2013)

Classes	Total
locatedInArea	46.864
origin	26.236
successor	567
deathOrBurialPlace	7.146
other	10.656
keyPerson	392
partOf	5.520
influencedBy	154
partner	190
parent	298
	98.023

Entretanto, durante o processamento do conjunto de dados disponibilizado por (BATISTA D.S.; FORTE, 2013) notou-se que existem 35 sentenças à mais no conjunto de dados disponibilizado pelos autores do que no conjunto de dados referenciado pelo trabalho dos mesmos. Os autores foram questionados sobre esta divergência, mas não souberam indicar com precisão o motivo da ocorrência da mesma.

O conjunto de dados disponibilizado pelos autores conta com **98.023** sentenças, **35** sentenças à mais que o apresentado pelos autores em seu trabalho [Tabela 1]. A distribuição de sentenças em cada classe no conjunto de dados disponibilizado por (BATISTA D.S.; FORTE, 2013) pode ser encontrado na Tabela 2.

Cada sentença do conjunto de dados é composta por uma sequência de palavras p_1, p_2, \dots, p_n de comprimento **n** e consta com algumas informações identificadas, essas são:

- A sentença extraída da *Wikipédia*
- Um campo informando se a classificação foi checada manualmente pelos autores

Tabela 3 – Distribuição de sentenças em classes no conjunto de dados gerado neste trabalho

Classes	Total
locatedInArea	44.408
origin	24.151
successor	564
deathOrBurialPlace	6.951
other	10.133
keyPerson	384
partOf	5.404
influencedBy	154
partner	190
parent	298
	92.637

- A primeira entidade e_1
- A segunda entidade e_2
- O tipo de relação entre as entidades

Neste trabalho, um novo conjunto de dados foi gerado a partir do processamento do conjunto de dados disponibilizado por (BATISTA D.S.; FORTE, 2013). Este novo conjunto de dados inclui apenas as sentenças cujo as entidades, ou pelo menos o primeiro token de cada entidade tokenizadas, aparecem na sentença, descartando assim sentenças que não contém as entidades marcadas. Com a aplicação deste critério, 5.386 sentenças foram descartadas. A quantidade de sentenças por classe no conjunto de dados gerado pode ser conferido na Tabela 3:

Cada entrada no dataset está representada seguindo o formato abaixo. Seste a sentença está contida no campo "SENTENCE", as entidades estão contidas nos campos "ENTITY1" e "ENTITY2", os campos "TYPE1" e "TYPE2" representam os respectivos tipos das entidades e "REL TYPE" representa o tipo da relação. As entradas ainda contam com um campo para representar se foram checadas manualmente pelos autores do dataset, este campo é o "MANUALLY CHECKED".

SENTENCE : A América Latina localiza-se totalmente no hemisfério ocidental, sendo a

MANUALLY CHECKED : TRUE

ENTITY1 : América Latina

TYPE1 : LOCATION

ENTITY2 : Argentina

TYPE2 : LOCATION

REL TYPE : locatedInArea

3.2 Word embeddings

Neste trabalho optou-se por utilizar vetores de palavras provenientes do repositório de vetores de palavras do *NLCI-Embeddings* (HARTMANN N.; FONSECA, 2017). Este repositório contém vetores de palavras formados a partir de um grande corpus com mais de 1 bilhão de tokens do português brasileiro e português europeu de fontes e gêneros variados. O treinamento dos vetores de palavra disponibilizados no repositório do NLCI-USP (Núcleo Interinstitucional de Linguística Computacional, USP) ocorreu com os algoritmos *Glove*, *Word2Vec*, *Wang2Vec* e *FastText*. Os modelos gerados a partir dos algoritmos *Word2vec*, *Wang2Vec* e *FastText* foram treinados utilizando os métodos de treinamento *CBOW* e *SkipGram*. Para cada um dos algoritmos foram gerados *embeddings* com 50, 100, 300, 600 e 1000 dimensões.

A partir dos *word embeddings* adotados foram removidos todos os *embeddings* cujo a dimensão divergia do valor determinado, ex.: Ao avaliar o conjunto de *embeddings FastText SkipGram* com 50 dimensões, foram descartados todos os *embeddings* que não tinham 50 dimensões.

Os experimentos realizados neste trabalho foram replicados utilizando vetores de palavras gerados a partir dos quatro algoritmos, considerando ambas as formas de treinamento disponível e foram utilizados vetores de 50 dimensões, 300 dimensões e 1000 dimensões.

3.3 Métricas de avaliação

Nas tarefas de extração de relações normalmente são utilizadas as métricas precisão, cobertura (*recall*) e medida F1. A acurácia é comumente desconsiderada pois em atividades que envolvem a classificação de dados desbalanceados esta medida acaba refletindo principalmente os resultados da classificação dos grupos majoritários. Assim, para cada relação da classe C as seguintes medidas foram levantadas:

- **Positivos verdadeiros (T_p):** Número de relações da classe C que foram classificadas corretamente pelo modelo;
- **Falsos positivos (F_p):** Número de relações de outras classes que foram classificados como pertencentes a classe C pelo modelo;
- **Falsos negativos (F_n):** Número de relações da classe C que foram classificadas como pertencentes a outras classes pelo modelo.

Precisão A precisão (P) representa a porcentagem de relações corretamente classificadas sobre o total de relações classificadas e é definida como:

$$P = \frac{T_p}{T_p + F_p} \quad (3.1)$$

Cobertura A cobertura (C) representa a porcentagem de relações corretamente classificadas sobre o total de relações presentes nos dados de teste:

$$C = \frac{T_p}{T_p + F_n} \quad (3.2)$$

Medida F1 A medida F1 é dada a partir da média harmônica entre a precisão e a cobertura, considerando-se importâncias equivalentes para as duas:

$$F_1 = \frac{2 \cdot P \cdot C}{P + C} \quad (3.3)$$

3.4 Arquitetura da rede neural convolucional

Optou-se por utilizar neste trabalho a arquitetura de rede neural convolucional (CNN) proposta por (MELO, 2018), ilustrada na Figura 3. Esta arquitetura engloba a representação vetorial de palavras, uma camada convolucional, uma camada de agrupamento e um classificador *softmax*.

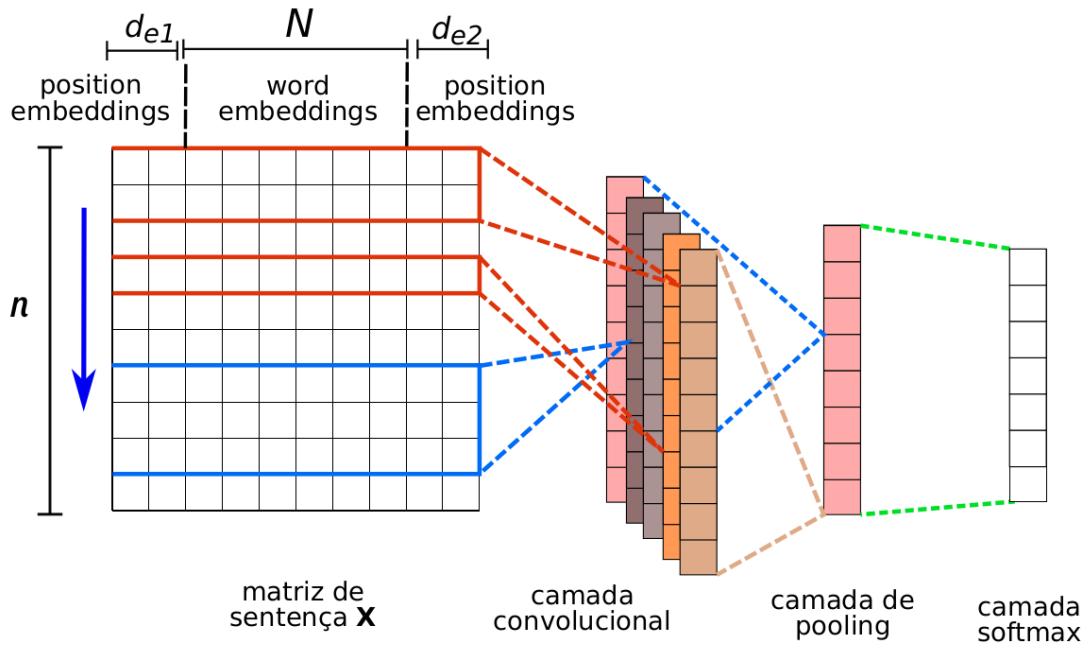


Figura 3 – Arquitetura de uma Rede Neural Convolucional para extração de relações

Fonte: (KIM, 2014)

A rede recebe como entrada uma matriz de sentenças \mathbf{X} com dimensão $n \cdot (N + 2 \cdot d_e)$, sendo cada sentença representada como $w_1, w_2, w_3, \dots, w_N$, onde cada w_i corresponde a uma palavra da sentença em sua codificação vetorial de N dimensões. Os vetores com a posição das entidades (*position embeddings*) de comprimento d_e também entram na conta, sendo representados pelo termo $2d_e$. Para formar a matriz de entrada, o comprimento de cada sentença

foi fixado em n . Assim, sentenças com menos de n palavras tiveram as posições restantes do vetor preenchidas com zero e sentenças com mais de n palavras foram truncadas, tendo sido removidas todas as palavras após a n -ésima palavra. A matriz de entrada é computada na camada convolucional, onde ocorre o aprendizado da rede. Diversas características são extraídas a partir das convoluções dos filtros com a matriz de entrada gerando os mapas de características. A última camada da rede entrega como saída um vetor de dimensão igual ao número de tipos de relações. Cada posição no vetor indica a probabilidade da sentença S pertencer a classe C , para toda sentença S na matriz de entrada e toda classe C pertencente ao domínio de classes do conjunto de dados. Essa probabilidade é representada por $p(S|C, \theta)$, sendo θ os parâmetros da rede.

3.4.1 A camada de entrada

Assim como em (MELO, 2018; NGUYEN T. H.; GRISHMAN, 2015; ZENG D.; LIU, 2014), a matriz de entrada da rede é composta pelas n primeiras palavras de cada sentença codificadas na forma vetorial e pelos vetores de posições (*position embeddings*). Nesta camada são projetadas as representações de palavras discretas x_t para vetores de baixa dimensão e_t , esse processo é descrito pela equação:

$$e_t = W_e \cdot X_t \quad (3.4)$$

Onde X_t denota uma representação *one-hot* de dimensão V da t -ésima palavra do vocabulário de comprimento V , W_e representa cada matriz de pesos pré-treinada a partir dos vetores de palavra adotados neste trabalho. A abordagem adotada propõe utilizar *position embeddings* sob a hipótese de que em sentenças mais longas as palavras mais distantes das entidades identificadas têm menor relevância para classificação do tipo da relação entre as entidades do que as palavras mais próximas. Para isso, os embeddings de posição contribuem com a posição relativa das palavras às entidades identificadas, tendo sido cada valor discreto de distância mapeado para um valor real. Visando diminuir a dimensionalidade de características dos *position embeddings* (MELO, 2018) propôs limitar a distância máxima e a distância mínima entre as palavras e as entidades. Em seu trabalho, é utilizado o valor de 40 unidades de distância como valor máximo e também como valor mínimo de distância. Esse valor também foi adotado neste trabalho. Esta forma de representação é chamada de *rectified position embeddings* e é determinada pelas equações:

$$d_{e_1} = \max(d_{\text{máx}}, i - i_{e_1}) \quad (3.5)$$

$$d_{e_2} = \max(-d_{\text{máx}}, i - i_{e_2}) \quad (3.6)$$

Sendo $d_{\text{máx}}$ o valor máximo de distância entre as entidades e as palavras, determinado como sendo 40 unidades de distância. Ao final desse processo, uma sentença \mathbf{S} estará representado como uma matriz \mathbf{M} de dimensão $n \cdot (N + 2 \cdot d_e)$, onde n é a quantidade de palavras na sentença, d_e é a dimensionalidade do *position embeddings* e N é a dimensionalidade dos *word embeddings*. Assim como em (MELO, 2018), neste trabalho optou-se por fixar $n = 50$ e $d_e = 50$. Já a dimensionalidade dos *embeddings* (N) variou entre 50, 300 e 1000 de acordo com os experimentos realizados.

3.4.2 A Camada convolucional

Após a camada de entrada a matriz de entrada \mathbf{X} passa para a camada convolucional para extrair as características. Dado um filtro (*kernel*) \mathbf{W} de comprimento k , sendo o filtro uma matriz de pesos $\mathbf{W} = [w_1, w_2, \dots, w_k]$ (onde cada w_i é vetor de dimensão $N + 2 \cdot d_e$). O objetivo desta camada é obter uma matriz de saída $\mathbf{O} = [o_1, o_2, \dots, o_k] = f(\mathbf{X} * \mathbf{W})$, onde cada o_i é representado por:

$$o_i = f\left(\sum_{a=0}^{k-1} w_a \cdot x_{i-a} + b\right) \quad (3.7)$$

onde w_a representa o vetor de pesos que conecta o neurônio da camada de convolução com os neurônios da camada de *pooling*, b é o escalar de polarização (*bias*) da camada e $f(\cdot)$ é a função de ativação.

Assim como em (MELO, 2018), neste trabalho foram utilizados 100 filtros de tamanho 3 e a função de ativação adotada foi linear retificadora, dada por:

$$f(x) = \max(0, x) \quad (3.8)$$

3.4.3 A camada de pooling

O vetor de saída da camada de convolução vai para a camada de *pooling*, que tem como objetivo agrupar as características presentes no mapa de características gerado como saída da camada de convolução e diminuir a dimensionalidade do mesmo. Para atingir esse objetivo aplicou-se o algoritmo *global pooling*, o resultado obtido com essa operação é dado por:

$$\begin{bmatrix} o_1 \\ o_2 \\ o_3 \\ \dots \\ o_k \end{bmatrix} \xrightarrow{\text{global pooling}} \begin{bmatrix} \max(o_1) \\ \max(o_2) \\ \max(o_3) \\ \dots \\ \max(o_k) \end{bmatrix} \quad (3.9)$$

Portanto, a saída da camada de *pooling* é dada por um vetor de comprimento k com o maior valor encontrado em cada mapa de característica o_i recebido da camada de convolução.

3.4.4 A camada de classificação

A saída resultante da camada de *pooling* é aplicado como entrada da camada de classificação. O objetivo desta camada é definir probabilidades para cada classe com base no vetor de características resultantes da camada anterior. A probabilidade total resultante deve somar 1.0 e é dada a partir da aplicação da função de *softmax*, dada por:

$$p_j = \frac{e^{z_j}}{\sum_{i=1}^k e^{z_i}} \quad (3.10)$$

$$f(X) = \arg_j \max p_j(X) \quad (3.11)$$

Portanto, em (3.10) o vetor de características tem seus valores normalizados em $[0, 1]$ e após essa operação é aplicada a função (3.11) que retorna a classe de maior probabilidade.

3.5 Experimentos e reprodutibilidade

Foram realizados 21 experimentos diferentes neste trabalho. Houveram variações no *word embedding* utilizado como entrada em cada experimento, mas os parâmetros da rede se mantiveram fixos durante todo o trabalho. A principal motivação para esta escolha foram os bons resultados encontrados por (MELO, 2018). Assim, as principais contribuições deste trabalho incluem as variações nos resultados a partir das variações nos *word embeddings* e a validação do modelo de (MELO, 2018) para textos em português.

Desta forma, a rede neural utilizada neste trabalho foi treinada por duas épocas, em mini lotes de 64 amostras, com o algoritmo *ADAM* (*Adaptative Moment Estimation*) com os parâmetros taxa de aprendizagem (θ): 0.001, β_1 : 0.9, β_2 : 0.999, decaimento: 0.0 e como função de custo foi adotada entropia cruzada.

Na camada de entrada foram utilizados os *word embeddings* *Word2Vec*, *Wang2Vec* e *FastText* treinados utilizando *Skip-gram* e *CBoW* com as dimensões de 50, 300 e 1000 e o *word embedding* *GloVe* também com as dimensões 50, 300 e 1000. O tamanho das sentenças foi fixado em 50 palavras e foram utilizados *position embeddings* de dimensão 2 x 50.

Na camada de convolução foram utilizados filtros de dimensão 100 x 3; sendo seguida por uma camada de agrupamento do tipo *global pooling* e de uma camada totalmente conectada *softmax* de comprimento igual ao número de classes de relações.

Para o treinamento do modelo foi utilizada validação cruzada *k-fold* com $k = 10$ e a qualidade global de classificação e a qualidade por classe de relação foram avaliadas utilizando as medidas de precisão, cobertura e métrica F1, conforme definidas na Seção 3.3 deste trabalho.

Como o conjunto de dados adotado neste trabalho não descreve explicitamente a direção do relacionamento em cada sentença, todas as relações neste trabalho foram adotadas como bi-direcionais, mantendo assim 10 classes possíveis para cada relação.

4

Experimentos, resultados e discussão

Neste capítulo são descritos os experimentos realizados neste trabalho, assim como são apresentados e discutidos os resultados encontrados.

Os experimentos foram iniciados utilizando o Sistema Computacional em Cluster do Laboratório de Alto Desempenho da Universidade Federal de Sergipe (LCAD) ¹. Entretanto, durante o desenvolvimento deste trabalho, problemas técnicos no Cluster o tornaram indisponível para continuação dos trabalhos realizados no LCAD.

Com o Cluster do LCAD indisponível, os experimentos descritos nesta seção foram finalizados utilizando o *Google Cloud Computer Engine* ², sistema de máquinas virtuais escalonáveis de alto desempenho disponibilizado pela Google Cloud Platform (GCP).

A GCP disponibiliza um período de testes de 12 meses com verba de \$300 dólares para serem gastos na plataforma a novos usuários ³. Os experimentos realizados neste trabalho utilizaram o período de testes do GCP e, portanto, não houveram custos financeiros decorrente ao uso da plataforma pelos autores deste trabalho.

A rede neural descrita na Seção 3.4 foi treinada utilizando os parâmetros descritos na Seção 3.5 e experimentos foram realizados utilizando o conjunto de dados descrito na Seção 3.1 e os *word embeddings* descritos na Seção 3.2. Os resultados da classificação de relações utilizando cada conjunto de *word embeddings* será apresentado e discutido nesta seção.

Apesar de não terem sido encontrados outros trabalhos que utilizem variações do conjunto de dados adotado neste trabalho e redes neurais convolucionais, os resultados encontrados podem ser comparados com os resultados apresentados pelo próprio Batista em (BATISTA D.S.; FORTE, 2013; BATISTA, 2016).

¹ <lcad.ufs.br>

² <<https://cloud.google.com/compute/>>

³ <<https://cloud.google.com/free/docs/gcp-free-tier>>

Tabela 4 – Classificação de relações utilizando o modelo de *word embeddings GloVe*

Classes	GloVe								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.94	0.95	0.94	0.96	0.94	0.95	0.93	0.97	0.95
origin	0.90	0.90	0.90	0.86	0.93	0.89	0.92	0.84	0.88
successor	0.50	0.60	0.55	0.52	0.56	0.54	0.64	0.60	0.62
deathOrBurialPlace	0.84	0.78	0.81	0.87	0.69	0.77	0.75	0.83	0.79
other	0.82	0.81	0.82	0.83	0.85	0.84	0.86	0.83	0.84
keyPerson	0.60	0.34	0.43	0.74	0.43	0.54	0.70	0.43	0.53
partOf	0.87	0.87	0.87	0.87	0.87	0.87	0.88	0.90	0.89
influencedBy	0.38	0.20	0.26	0.54	0.48	0.51	0.59	0.43	0.50
partner	0.66	0.44	0.53	0.66	0.33	0.44	0.65	0.61	0.63
parent	0.60	0.69	0.64	0.70	0.68	0.69	0.69	0.61	0.65
média/total	0.71	0.66	0.67	0.75	0.68	0.70	0.76	0.70	0.73

A abordagem adotada por (BATISTA D.S.; FORTE, 2013; BATISTA, 2016) utiliza comparação de assinaturas *min-hash* (BRODER, 1997; BRODER A.; CHARIKAR, 2000) e *kNN* para classificar relações semânticas entre entidades. Ao comparar os resultados encontrados neste trabalho com os resultados apresentados em (BATISTA D.S.; FORTE, 2013; BATISTA, 2016), nota-se que o modelo apresentado neste trabalho obteve um desempenho particularmente interessante.

Na tabela 4 são apresentados os resultados obtidos a partir da classificação das relações utilizando o modelo de *word embeddings GloVe* nas três dimensões. A partir destes resultados pode-se notar uma constante variação da métrica F1 ao variar a quantia de dimensões entre 50, 300 e 1000.

Nas tabelas 5 e 6 são apresentados os resultados obtidos a partir da classificação das relações utilizando o modelo de *word embeddings Word2Vec* utilizando os métodos de treinamento *CBOW* e *SkipGram* e nas três dimensões.

Os resultados mostram que para 50 dimensões há uma baixa variação nos resultados encontrados a partir dos diferentes métodos de treinamento. Entretanto, ao aumentar a quantidade de dimensões também percebe-se um aumento na variação entre estes resultados. Percebeu-se também que quando utilizado o método de treinamento *CBOW* a precisão obtida foi mais alta que ao utilizar o método *SkipGram*. Entretanto o contrário aconteceu à cobertura, isto é, foram obtidos resultados mais altos de cobertura utilizando o método *CBOW*.

As tabelas 7 e 8, por sua vez, apresentam os resultados obtidos a partir da classificação das relações utilizando o modelo de *word embeddings Wang2Vec* utilizando os métodos de treinamento *CBOW* e *SkipGram* e nas três dimensões.

Comparando as tabelas 7 e 8 pode-se notar que ao treinar a rede utilizando o modelo de *word embeddings SkipGram* os resultados alcançados são notavelmente melhores que os

Tabela 5 – Classificação de relações utilizando o modelo de *word embeddings* Word2Vec CBoW

Classes	Word2Vec - CBoW								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.93	0.94	0.93	0.95	0.95	0.95	0.94	0.96	0.95
origin	0.88	0.87	0.88	0.88	0.90	0.89	0.90	0.89	0.90
successor	0.44	0.47	0.45	0.53	0.59	0.56	0.66	0.54	0.60
deathOrBurialPlace	0.84	0.74	0.79	0.80	0.77	0.78	0.82	0.78	0.80
other	0.74	0.82	0.78	0.83	0.80	0.82	0.86	0.83	0.84
keyPerson	0.61	0.22	0.32	0.56	0.51	0.54	0.58	0.46	0.51
partOf	0.81	0.83	0.82	0.87	0.86	0.87	0.83	0.90	0.87
influencedBy	0.43	0.07	0.11	0.62	0.28	0.39	0.87	0.28	0.43
partner	0.69	0.19	0.30	0.63	0.33	0.44	0.75	0.32	0.44
parent	0.57	0.56	0.56	0.74	0.73	0.74	0.69	0.61	0.65
média/total	0.69	0.57	0.59	0.74	0.67	0.70	0.79	0.66	0.70

Tabela 6 – Classificação de relações utilizando o modelo de *word embeddings* Word2Vec Skip-Gram

Classes	Word2Vec - Skip-Gram								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.91	0.96	0.94	0.96	0.94	0.95	0.94	0.97	0.95
origin	0.90	0.87	0.89	0.89	0.91	0.90	0.89	0.92	0.90
successor	0.57	0.44	0.49	0.59	0.54	0.56	0.68	0.55	0.61
deathOrBurialPlace	0.80	0.79	0.80	0.80	0.80	0.80	0.87	0.76	0.81
other	0.85	0.75	0.79	0.81	0.87	0.84	0.88	0.80	0.84
keyPerson	0.61	0.33	0.43	0.59	0.52	0.55	0.62	0.52	0.57
partOf	0.83	0.86	0.85	0.92	0.83	0.87	0.91	0.87	0.89
influencedBy	0.00	0.00	0.00	0.65	0.52	0.58	0.46	0.48	0.47
partner	0.44	0.19	0.27	0.46	0.58	0.51	0.68	0.53	0.59
parent	0.57	0.58	0.57	0.68	0.66	0.67	0.68	0.70	0.69
média/total	0.65	0.58	0.60	0.73	0.72	0.72	0.76	0.71	0.73

resultados utilizando o modelo de *word embeddings* CBoW. Fica claro também uma baixa variação nos resultados ao aumentar a quantia de dimensões dos *word embeddings* de 300 para 1000 dimensões, diferente do que ocorre ao aumentar a quantia de dimensões de 50 para 300.

Acredita-se que a característica de considerar a ordem das palavras do Wang2Vec, tal qual a capacidade de prever o contexto do modelo treinado com *Skip-Gram*, contribuíram para o bom resultado obtido pelo modelo nos experimentos realizados. Comparado com o *Word2Vec*, tabelas 5 e 6, fica notória a melhora nos resultados alcançados.

As tabelas 7 e 8, por sua vez, apresentam os resultados obtidos a partir da classificação das relações utilizando o modelo de *word embeddings* Wang2Vec utilizando os métodos de treinamento CBOW e SkipGram e nas três dimensões.

Comparando as tabelas 7 e 8 pode-se notar que ao treinar a rede utilizando o modelo

Tabela 7 – Classificação de relações utilizando o modelo de *word embeddings* Wang2Vec CBoW

Classes	Wang2Vec - CBoW								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.96	0.93	0.94	0.96	0.95	0.96	0.93	0.96	0.94
origin	0.88	0.91	0.89	0.89	0.92	0.90	0.89	0.88	0.89
successor	0.55	0.60	0.58	0.66	0.50	0.57	0.70	0.53	0.60
deathOrBurialPlace	0.85	0.75	0.80	0.83	0.80	0.81	0.81	0.78	0.79
other	0.74	0.88	0.80	0.83	0.87	0.85	0.83	0.81	0.82
keyPerson	0.62	0.31	0.42	0.72	0.46	0.56	0.42	0.55	0.47
partOf	0.88	0.85	0.86	0.88	0.90	0.89	0.88	0.84	0.86
influencedBy	0.43	0.26	0.32	0.56	0.30	0.39	0.79	0.33	0.46
partner	0.58	0.26	0.36	0.72	0.54	0.62	0.68	0.33	0.45
parent	0.57	0.64	0.61	0.70	0.69	0.70	0.70	0.61	0.65
média/total	0.71	0.64	0.66	0.77	0.69	0.72	0.76	0.66	0.69

Tabela 8 – Classificação de relações utilizando o modelo de *word embeddings* Wang2Vec Skip-Gram

Classes	Wang2Vec - Skip-Gram								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.94	0.96	0.95	0.96	0.95	0.96	0.97	0.95	0.96
origin	0.90	0.90	0.90	0.91	0.90	0.90	0.90	0.91	0.90
successor	0.60	0.56	0.58	0.69	0.59	0.63	0.72	0.58	0.64
deathOrBurialPlace	0.84	0.80	0.82	0.82	0.81	0.82	0.78	0.84	0.81
other	0.86	0.81	0.83	0.81	0.87	0.84	0.86	0.87	0.86
keyPerson	0.62	0.55	0.58	0.75	0.39	0.51	0.71	0.55	0.62
partOf	0.89	0.88	0.88	0.87	0.92	0.89	0.90	0.89	0.89
influencedBy	1.00	0.02	0.04	0.70	0.35	0.46	0.63	0.37	0.47
partner	0.48	0.60	0.53	0.80	0.63	0.71	0.72	0.63	0.67
parent	0.50	0.66	0.56	0.68	0.77	0.72	0.82	0.62	0.71
média/total	0.76	0.67	0.68	0.80	0.73	0.74	0.80	0.72	0.75

de *word embeddings* SkipGram os resultados alcançados são notavelmente melhores que os resultados utilizando o modelo de *word embeddings* CBoW. Fica claro também uma baixa variação nos resultados ao aumentar a quantia de dimensões dos *word embeddings* de 300 para 1000 dimensões, diferente do que ocorre ao aumentar a quantia de dimensões de 50 para 300.

Acredita-se que a característica de considerar a ordem das palavras do Wang2Vec, tal qual a capacidade de prever o contexto do modelo treinado com Skip-Gram, contribuíram para o bom resultado obtido pelo modelo nos experimentos realizados. Comparado com o Word2Vec, tabelas 5 e 6, fica notória a melhora nos resultados alcançados.

Os últimos experimentos realizados utilizaram o modelo de *word embeddings* FastText. Os resultados obtidos a partir deste são apresentados nas 9 e 10.

Mesmo este sendo um modelo simples e de baixa profundidade (tendo em vista que

Tabela 9 – Classificação de relações utilizando o modelo de *word embeddings* *FastText CBoW*

Classes	FastText - CBoW								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.93	0.96	0.94	0.94	0.96	0.95	0.95	0.96	0.95
origin	0.91	0.86	0.88	0.90	0.90	0.90	0.90	0.91	0.90
successor	0.49	0.39	0.44	0.53	0.46	0.49	0.45	0.66	0.53
deathOrBurialPlace	0.80	0.78	0.79	0.82	0.78	0.80	0.83	0.79	0.81
other	0.80	0.81	0.81	0.87	0.80	0.83	0.85	0.84	0.84
keyPerson	0.65	0.26	0.37	0.59	0.28	0.38	0.58	0.37	0.46
partOf	0.86	0.87	0.86	0.84	0.89	0.87	0.88	0.88	0.88
influencedBy	0.17	0.09	0.12	0.58	0.33	0.42	0.64	0.39	0.49
partner	0.44	0.37	0.40	0.73	0.47	0.57	0.59	0.47	0.52
parent	0.47	0.69	0.56	0.50	0.76	0.60	0.66	0.59	0.62
média/total	0.65	0.61	0.62	0.73	0.66	0.68	0.73	0.69	0.70

Tabela 10 – Classificação de relações utilizando o modelo de *word embeddings* *FastText SkipGram*

Classes	FastText - SkipGram								
	50 dimensões			300 dimensões			1000 dimensões		
	P	C	F1	P	C	F1	P	C	F1
locatedInArea	0.95	0.96	0.95	0.96	0.95	0.95	0.98	0.93	0.95
origin	0.89	0.92	0.91	0.92	0.90	0.91	0.90	0.92	0.91
successor	0.52	0.54	0.53	0.41	0.74	0.53	0.70	0.48	0.57
deathOrBurialPlace	0.83	0.77	0.80	0.82	0.82	0.82	0.83	0.81	0.82
other	0.84	0.81	0.83	0.79	0.89	0.84	0.75	0.92	0.83
keyPerson	0.59	0.38	0.47	0.61	0.46	0.52	0.54	0.37	0.44
partOf	0.88	0.89	0.88	0.92	0.88	0.90	0.92	0.88	0.90
influencedBy	0.57	0.09	0.15	0.51	0.41	0.46	0.66	0.41	0.51
partner	0.60	0.46	0.52	0.52	0.46	0.49	0.39	0.77	0.51
parent	0.72	0.57	0.63	0.72	0.51	0.60	0.80	0.70	0.75
média/total	0.74	0.64	0.67	0.72	0.70	0.70	0.75	0.72	0.72

o mesmo contém apenas uma camada convolucional, uma camada de pooling e uma camada totalmente conectada) pode-se notar que para a tarefa de extração de relações simétricas (onde não é considerada a direção da relação) os resultados obtidos neste trabalho são superiores aos resultados obtidos pelo autor do conjunto de dados quando comparadas as macrométricas e quando comparado o resultado por classes.

A Tabela 11 apresenta a comparação entre os resultados obtidos neste trabalho e os resultados obtidos em (BATISTA D.S.; FORTE, 2013; BATISTA, 2016) comparando apenas as macrométricas e a Tabela 12 apresenta a comparação por classe considerando o experimento realizado neste trabalho utilizando o algoritmo Wang2Vec SkipGram 1000 dimensões e os resultados apresentados por (BATISTA D.S.; FORTE, 2013; BATISTA, 2016).

Além do notável desempenho apresentado pelo modelo adotado, ao observar os resul-

Modelos	P	C	F1
Este trabalho - Wang2Vec SkipGram 1000 dimensões	0.80	0.72	0.75
Este trabalho - Wang2Vec SkipGram 300 dimensões	0.80	0.73	0.74
Este trabalho - Word2Vec SkipGram 1000 dimensões	0.76	0.71	0.73
Este trabalho - FastText SkipGram 1000 dimensões	0.75	0.72	0.72
Este trabalho - GloVe 1000 dimensões	0.76	0.70	0.73
Este trabalho - Wang2Vec CBoW 300 dimensões	0.77	0.69	0.72
Este trabalho - Word2Vec SkipGram 300 dimensões	0.73	0.72	0.72
Este trabalho - Word2Vec CBoW 1000 dimensões	0.79	0.66	0.70
Este trabalho - GloVe 300 dimensões	0.75	0.68	0.70
Este trabalho - FastText SkipGram 300 dimensões	0.72	0.70	0.70
Este trabalho - FastText CBoW 1000 dimensões	0.73	0.69	0.70
Este trabalho - Wang2Vec CBoW 1000 dimensões	0.76	0.66	0.69
Este trabalho - Word2Vec CBoW 300 dimensões	0.74	0.67	0.70
Este trabalho - FastText CBoW 300 dimensões	0.73	0.66	0.68
Este trabalho - FastText SkipGram 50 dimensões	0.74	0.64	0.67
Este trabalho - GloVe 50 dimensões	0.71	0.66	0.67
Este trabalho - Wang2Vec SkipGram 50 dimensões	0.71	0.64	0.66
Este trabalho - Wang2Vec CBoW 50 dimensões	0.71	0.64	0.66
Este trabalho - FastText CBoW 50 dimensões	0.65	0.61	0.62
Este trabalho - Word2Vec CBoW 50 dimensões	0.69	0.57	0.59
Este trabalho - Word2Vec SkipGram 50 dimensões	0.65	0.58	0.60
(BATISTA D.S.; FORTE, 2013 ; BATISTA, 2016)	0.736	0.443	0.553

Tabela 11 – Comparação entre macrométricas dos trabalhos que utilizam o mesmo conjunto de dados

Classe	Este trabalho			Batista 2013		
	P	C	F1	P	C	F1
locatedInArea	0.97	0.95	0.96	0.924	0.922	0.923
origin	0.90	0.91	0.91	0.733	0.908	0.811
sucessor	0.72	0.58	0.64	0.541	0.161	0.248
deathOrBurialPlace	0.78	0.84	0.81	0.800	0.574	0.671
other	0.86	0.87	0.86	0.767	0.53	0.63
keyPerson	0.71	0.55	0.62	0.233	0.079	0.117
partOf	0.90	0.89	0.89	0.680	0.576	0.623
influencedBy	0.63	0.37	0.47	0.000	0.000	0.000
partner	0.72	0.63	0.67	0.600	0.188	0.286
parent	0.82	0.62	0.71	0.545	0.727	0.623

Tabela 12 – Comparação entre resultados por classes: Este trabalho - Wang2Vec SkipGram 1000 dimensões e ([BATISTA D.S.; FORTE, 2013](#); [BATISTA, 2016](#))

tados obtidos ao final dos experimentos realizados pôde ser percebido que houve uma baixa variação (em média de ± 0.02) na performance geral (entende-se por performance geral a média ponderada das métricas definidas na Seção 3.1 deste trabalho) da *CNN* ao variar a dimensão dos *word embeddings* de 300 para 1000. Cenário bem diferente do encontrado ao variar a dimensão dos *word embeddings* de 50 para 300. Esta observação reforça os resultados encontrados por (MELAMUD O.; MCCLOSKEY, 2016; PATEL K.; BHATTACHARYYA, 2017) nos quais é mostrado que a contribuição do número de dimensões nos *word embeddings* para tarefas intrínsecas tem seu pico com uso de *word embeddings* com um número de dimensões próximo a 50 e para tarefas extrínsecas o pico se encontra próximo de 300 dimensões.

Pôde ser notado também que os modelos treinados com *word embeddings* treinados utilizando *Skip-Gram* obtiveram melhor desempenho que os treinados com *CBoW*. Acredita-se que esse pode ser um indicativo da importância da capacidade de prever o contexto dos *word embeddings* treinados com *Skip-Gram* para tarefas de classificação de relações semânticas.

Os resultados apresentados por este trabalho reforçam o uso de redes neurais convolucionais como ferramenta para extração de características de alto nível a partir de sentenças de textos codificados como *word embeddings* somados ao uso de *position embeddings*. Assim como realça a importância da quantidade e distribuição dos dados por classe no conjunto de treinamento, notou-se que classes como *partner* e *influencedBy* tiveram desempenho notoriamente inferior à média ponderada total. Estas classes em especial tinham apenas 190 e 154 amostras, respectivamente, ambas as quantidades de amostras menores que 1% do tamanho total do conjunto de dados.

5

Conclusão

A realização de tarefas de processamento de linguagem natural no idioma português, brasileiro ou português, apresenta diversos desafios. Um dos principais desafios enfrentados durante a realização deste trabalho foi a dificuldade em encontrar conjuntos de dados rotulados em português para uso em tarefas de extração de relações semânticas binárias. Acredita-se que o principal fator limitante para a produção desses conjuntos de dados é o enorme esforço humano necessário para rotular manualmente os dados. A revisão bibliográfica realizada neste trabalho aponta a abordagem de supervisão distante como bastante promissora para se desenvolver grandes conjuntos de dados com menor esforço necessário, tornando possível assim expandir a quantidade de conjuntos de dados rotulados para uso em tarefas de PLN.

O trabalho apresentado neste documento teve como principal objetivo a validação do modelo de Rede Neural Convolucional proposto por (MELO, 2018) para extração de relações semânticas binárias em dados não estruturados em português. Para isso, foi utilizado um conjunto de dados com 92.637 amostras divididas entre 10 relações semânticas gerados a partir da abordagem de supervisão distante por (BATISTA D.S.; FORTE, 2013). Os resultados encontrados apresentam bons valores de precisão, cobertura e média F1; sendo estes, em determinados casos, superiores aos encontrados pelos autores do conjunto de dados em seus experimentos.

Dentre os experimentos realizados neste trabalho, pôde ser notado que a quantidade de dimensões dos *word embeddings* utilizados para codificar as sentenças que alimentaram a rede neural convolucional não apresentavam uma influência linear nas macro-métricas produzidas na avaliação do modelo. Percebeu-se que nesta tarefa de extração de relações quanto utilizado *word embeddings* com dimensão superior a 300 para treinamento do modelo, o resultado final não sofre grande variação.

Notou-se que ao utilizar *word embeddings* com 1000 dimensões, houve variação de até 0.03% nas macro-métricas adotadas, o que reforça resultados já apresentados por outros autores na literatura indicando que em tarefas de aprendizagem de máquina extrínsecas a contribuição do

número de dimensões dos *word embeddings* para a qualidade do modelo (medida em função das macro-métricas adotadas) tem seu pico em cerca de 300 dimensões.

Muito ainda pode ser realizado na área de extração de relações semânticas em português. Pesquisas futuras, para este estudo apresentado, podem incluir a realização de experimentos em conjuntos de dados mais balanceados, uso de camadas profundas no modelo adotado neste trabalho e modificações no modelo adotado para permitir a extração de relações semânticas não binárias. O código fonte utilizado neste trabalho está disponível de forma aberta na plataforma GitHub e pode ser acessado em <https://github.com/arielrodrigues/nlp-relations-extraction-ptbr>.

Os métodos desenvolvidos neste trabalho estão sendo aplicados a um *pipeline* de identificação e extração de promessas dos candidatos à Chefia do Executivo brasileiro, através do processamento dos seus planos de governo. A implementação deste *pipeline* está em andamento e os resultados alcançados serão disponibilizados como parte da Plataforma LADDRES (PASSOS, 2018).

Referências

- ALVA-MANCHEGO F. E.; ROSA, J. L. G. Towards semi-supervised brazilian portuguese semantic role labeling: Building a benchmark. *H. Caseli, A. Villavicencio, A. Teixeira, and F. Perdigão (Eds.), Computational Processing of the Portuguese Language. Lecture Notes in Computer Science*, v. 7243, p. 210–217, 2012. Springer Berlin Heidelberg 2012b. Citado na página 18.
- ANDERSEN, P. M. e. a. Automatic extraction of facts from press releases to generate news stories. *Proceedings of the 3rd Conference on Applied Natural Language Processing*, p. 170–177, abril 1992. Citado na página 9.
- BANKO M.; CAFARELLA, M. J. S. S. B. M. E. O. Open information extraction from the web. *IJCAI*, v. 7, p. 2670–2676, 2007. Citado na página 12.
- BARONI M.; DINU, G. K. G. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, I, p. 238–247, 2014. Citado na página 13.
- BATISTA, D. S. *Large-Scale Semantic Relationship Extraction for Information Discovery*. Tese (Doutorado) — Instituto Superior Técnico - Universidade de Lisboa, junho 2016. Citado 5 vezes nas páginas 4, 29, 30, 33 e 34.
- BATISTA D.S.; FORTE, D. S. R. M. B. S. M. Extração de relações semânticas de textos em português explorando a dbpédia e a wikipédia. *Linguamatica 5(1)*, p. 41–57, 2013. Citado 10 vezes nas páginas 4, 18, 20, 21, 22, 29, 30, 33, 34 e 36.
- BENGIO Y.; DUCHARME, R. V. P. J. C. A neural probabilistic language model. *Journal of Machine Learning Research*, v. 3, p. 1137–1155, 2003. Citado na página 13.
- BICK, E. Automatic semantic role annotation for portuguese. *Proceedings of TIL 2007 - 5th Workshop on Information and Human Language Technology / Anais do XXVII Congresso da SBC*, p. 1713–1716, 2007. Citado na página 18.
- BOJANOWSKI P.; GRAVE, E. J. A. M. T. *Enriching Word Vectors with Subword Information*. [S.l.], 2016. ArXiv preprint arXiv:1607.04606. Citado na página 14.
- BOLLACKER K.; EVANS, C. P. P. S. T. T. J. Freebase: a collaboratively created graph database for structuring human knowledge. *SIGMOD '08*, p. 1247–1250, 2008. Citado na página 12.
- BRODER, A. On the resemblance and containment of documents. *Proceedings of the Conference on Compression and Complexity of Sequences*, 1997. Citado na página 30.
- BRODER A.; CHARIKAR, M. F. A. M. M. M. Min-wise independent permutations. *Journal of Computer and System Sciences*, 2000. 60(3). Citado na página 30.
- CARDOSO, N. Rembrandt - reconhecimento de entidades mencionadas baseado em relações e análise detalhada do texto. *Linguatca*, p. 195–211, 2008. Citado na página 18.

COLLOBERT R.; WESTON, J. B. L. K. M. K. K. P. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, v. 12, p. 2493–2537, 2011. Citado 2 vezes nas páginas 13 e 15.

CRAVEN M.; KUMLIEN, J. Constructing biological knowledge bases by extracting information from text sources. *ISMB*, v. 1999, p. 77–78, 1999. Citado na página 12.

CRUZ G. A.; WEITZEL, L. Evaluation of relation extraction systems for portuguese language pt-br. p. 1–6, 2018. Citado na página 18.

DODDINGTON, G. e. a. The automatic content extraction (ace) program—tasks, data, and evaluation. *Proceedings of LREC*, p. 837–840, 2004. Citado na página 10.

DUMAIS S. T.; FURNAS, G. W. L. T. K. D. S. H. R. Using latent semantic analysis to improve access to textual information. *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, p. 281–285, 1988. Citado na página 13.

FADER A.; SODERLAND, S. E. O. Identifying relations for open information extraction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (Association for Computational Linguistics, 2011)*, p. 1535–1545, 2011. Citado na página 18.

FONSECA E. R.; ROSA, J. L. An architecture for semantic role labeling on portuguese. *H. Caseli, A. Villavicencio, A. Teixeira, and F. Perdigão (Eds.), PROPOR 2012. LNCS/LNAI*, v. 7243, p. 204–209, 2012. Springer Berlin Heidelberg 2012a. Citado na página 18.

FONSECA E. R.; ROSA, J. L. A two-step convolutional neural network approach for semantic role labeling. *IEEE International Joint Conference on Neural Networks*, p. 2955–2961, 2013. Dallas, USA. Citado na página 18.

FREITAS M. C. DE; QUENTAL, V. Subsídios para a elaboração automática de taxonomia. *Anais do XXVII Congresso da SBC (V Workshop em Tecnologia da Informação e da Linguagem Humana TIL)*, p. 1585–1594, 2007. Citado na página 18.

GRANGER, R. H. Foul-up: A program that figures out meanings of words from context. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, p. 172–178, 1977. Citado na página 9.

HARTMANN N.; FONSECA, E. S. C. T. M. R. J. A. S. *Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks*. [S.l.], 2017. ArXiv:1708.06025. Citado 2 vezes nas páginas 13 e 23.

HARTMANN N. S.; DURAN, M. S. A. S. M. Automatic semantic role labeling on non-revised syntactic trees of journalistic texts. 2017. ArXiv:1704.03016. Citado na página 18.

HEARST, M. Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th Conference on Computational Linguistics*, pp. 539–545. Association for Computational Linguistics, 1992. Citado na página 11.

HOFFMANN R.; ZHANG, C. L. X. Z. L.; WELD, D. S. Knowledge-based weak supervision for information extraction of overlapping relations. *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, HLT’11*, 2011. Citado na página 13.

- HUMPHREYS K., G. R. A. S. H. C. M. B. C. H. W. Y. University of sheffield: description of the lasie-ii system as used for muc-7. *Proceedings of MUC-7*, 1998. Citado na página 11.
- JOHNSON R.; ZHANG, T. *Effective use of word order for text categorization with convolutional neural networks*. [S.l.], 2014. ArXiv preprint arXiv:1412.1058. Citado na página 16.
- JURAFSKY D.; MARTIN, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. [S.l.]: Prentice-Hall, 2009. Citado na página 9.
- KALCHBRENNER N.; GREFFENSTETTE, E. B. P. *A convolutional neural network for modelling sentence*. [S.l.], 2014. ArXiv preprint arXiv:1404.2188. Citado na página 16.
- KIM, Y. *Convolutional neural networks for sentence classification*. [S.l.], 2014. ArXiv preprint arXiv:1408.5882. Citado 2 vezes nas páginas 16 e 24.
- LAI S.; XU, L. L. K. Z. J. Recurrent convolutional neural networks for text classification. *AAAI*, v. 333, p. 2267–2273, 2015. Citado 2 vezes nas páginas 13 e 15.
- LECUN Y.; BOTTOU, L. B. Y. H. P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, novembro 1998. 86(11):22782324. Citado na página 15.
- LEHNERT, W. Human and computational question answering. *Cognitive Science*, p. 47–73, janeiro 1977. Citado na página 9.
- LING W.; DYER, C. B. A. T. I. Two/too simple adaptations of word2vec for syntax problems. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics.*, 2015. Citado 2 vezes nas páginas 13 e 14.
- LUND K.; BURGESS, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, Computers*, v. 28, p. 203–208, 1996. Citado na página 13.
- MELAMUD O.; MCCLOSKEY, D. P. S. B. M. *The Role of Context Types and Dimensionality in Learning Word Embeddings*. [S.l.], 2016. ArXiv:1601.00893. Citado na página 35.
- MELO, F. S. *Extração de relações a partir de dados não estruturados baseada em deep learning e supervisão distante*. Dissertação (Mestrado) — Departamento de Computação, Universidade Federal de Sergipe, 2018. Citado 11 vezes nas páginas 8, 12, 15, 16, 17, 20, 24, 25, 26, 27 e 36.
- MESSAGE understanding conference-6: A brief history. *Proceedings of the 16th International Conference on Computational Linguistics*, p. 466 – 471, 1996. Citado na página 10.
- MIKOLOV T.; CHEN, K. C. G. D. J. Efficient estimation of word representations in vector space. *Proceedings of International Conference on Learning Representations Workshop (ICLR-2013)*, 2013. Citado 2 vezes nas páginas 13 e 14.
- MINTZ M.; BILLS, S. S. R. J. D. Distant supervision for relation extraction without labeled data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, p. 1003–1011, 1992. Citado na página 12.

- MINTZ M.; BILLS, S. S. R. J. D. Distant supervision for relation extraction without labeled data. *Proceedings of the Annual Meeting of the Association for Computational Linguistics and of the International Conference on Natural Language*, 2009. Citado na página 12.
- NGUYEN T. H.; GRISHMAN, R. Relation extraction: Perspective from convolutional neural networks. *Proceedings of NAACL-HLT*, p. 39–48, 2015. [S.l.: s.n.]. Citado 3 vezes nas páginas 17, 18 e 25.
- NGUYEN, T. T.; MOSCHITTI, A. End-to-end relation extraction using distant supervision from external semantic repositories. *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers*, 2011. Citado na página 12.
- NIGAM K.; MCCALLUM, A. T. S. M. T. Learning to classify text from labeled and unlabeled documents. *AAAI/IAAI*, p. 792, 1998. Citado na página 12.
- OLIVEIRA H. G.; SANTOS, D. G. P. Extração de relações semânticas entre palavras a partir de um dicionário: o papel e sua avaliação. *Linguamática*, v. 2, n.1, p. 77–94, 2010. Citado na página 18.
- PASSOS, N. R. S. *Transformando dados em conhecimento: LADDRES, uma aplicação prática*. 62 p. Monografia (Engenharia de Computação) — Universidade Federal de Sergipe, Sergipe, 2018. Citado na página 37.
- PATEL K.; BHATTACHARYYA, P. Towards lower bounds on number of dimensions for word embeddings. *Proceedings of the The 8th International Joint Conference on Natural Language Processing*, p. 31–36, dezembro 2017. Citado na página 35.
- PENNINGTON J.; SOCHER, R. M. C. D. Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-2014)*, v. 12, p. 132–1543, 2014. Citado 2 vezes nas páginas 13 e 14.
- PEREIRA V.; PINHEIRO, V. Report - um sistema de extração de informações aberta para língua portuguesa. *Proceedings of Symposium in Information and Human Language Technology, SBC*, november 2015. Citado na página 18.
- RIEDEL S.; YAO, L.; MCCALLUM, A. Modeling relations and their mentions without labeled text. *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD'10*, 2010. Citado na página 13.
- RIESBECK C. K.; SCHANK, R. C. Comprehension by computer: Expectation-based analysis of sentences in context. *Technical report, DTIC Document*, 1976. Citado na página 9.
- RILOFF E.; JONES, R. Learning dictionaries for information extraction by multi-level bootstrapping. *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference (AAAI '99/IAAI '99)*, p. 474–479, 1999. Citado na página 11.
- SANTOS C. N. D.; XIANG, B. Z. B. *Classifying relations by ranking with convolutional neural networks*. [S.l.], 2015. ArXiv preprint arXiv:1504.06580. Citado na página 17.
- SCHANK, R. C. The primitive acts of conceptual dependency. *TINLAP-1*, janeiro 1975. Citado na página 9.

- SCHMITZ M.; BART, R. S. S. E. O. Open language learning for information extractio. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, p. 523–534, 2012. Citado na página 12.
- SEQUEIRA J.; GONÇALVES, T. Q. P. Semantic role labeling for portuguese – a preliminary approach. *H. Caseli, A. Villavicencio, A. Teixeira, and F. Perdigão (Eds.), PROPOR 2012. LNCS/LNAI*, v. 7243, p. 193–203, 2012. Springer Berlin Heidelberg 2012a. Citado na página 18.
- SILVA R. P. DA; SANTOS, F. A. N. F. B. d. M. H. T. Cross-language approach for sentiment classification in brazilian portuguese with convnets. *Information Technology-New Generations*, p. 311–316, 2018. [S.l.]: Springer. Citado na página 16.
- SUCHANEK F. M.; KASNECI, G.; WEIKUM, G. Yago: A core of semantic knowledge. *Proceedings of the 16th International Confer-ence on World Wide Web, WWW’07*, 2007. Citado na página 12.
- TAKAMATSU S.; SATO, I.; NAKAGAWA, H. Reducing wrong labels in distant supervision for relation extraction. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, ACL ’12*, 2012. Citado na página 13.
- XIANG Y.; CHEN, Q. W. X. Q. Y. Distant supervision for relation extraction with ranking-based methods. *Entropy n. 204*, v. 18, 2016. Citado na página 12.
- XU K.; FENG, Y. H. S. Z. D. *Semantic relation classification via convolutional neural networks with simple negative sampling*. [S.l.], 2015. ArXiv preprint arXiv:1506.07650. Citado na página 17.
- ZENG D.; LIU, K. C. Y. Z. J. Distant supervision for relation extraction via piecewise convolutional neural networks. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 17–21, 2015. [S.l.: s.n.]. Citado na página 17.
- ZENG D.; LIU, K. L. S. Z. G. Z. J. Relation classification via convolutional deep neural network. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, p. 2335–2344, 2014. Citado 3 vezes nas páginas 12, 17 e 25.
- ZHANG X.; LECUN, Y. *Text understanding from scratch*. [S.l.], 2015. ArXiv preprint arXiv:1502.01710. Citado na página 15.